

Tests multiples simulés et tests de normalité basés sur plusieurs moments dans les modèles de régression

Jean-Marie Dufour, Abdeljelil Farhat, Lynda Khalaf et Jean-Marie Dufour

Volume 80, numéro 2-3, juin–septembre 2004

Hommage à Marcel Dagenais

URI : <https://id.erudit.org/iderudit/011397ar>

DOI : <https://doi.org/10.7202/011397ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

HEC Montréal

ISSN

0001-771X (imprimé)

1710-3991 (numérique)

[Découvrir la revue](#)

Citer cet article

Dufour, J.-M., Farhat, A., Khalaf, L. & Dufour, J.-M. (2004). Tests multiples simulés et tests de normalité basés sur plusieurs moments dans les modèles de régression. *L'Actualité économique*, 80(2-3), 501–522.
<https://doi.org/10.7202/011397ar>

Résumé de l'article

Cet article illustre l'applicabilité des méthodes de rééchantillonnage dans le cadre des tests multiples (simultanés), pour divers problèmes économétriques. Les hypothèses simultanées sont une conséquence habituelle de la théorie économique, de sorte que le contrôle de la probabilité de rejet de combinaisons de tests est un problème que l'on rencontre fréquemment dans divers contextes économétriques et statistiques. À ce sujet, on sait que le fait d'ignorer le caractère conjoint des hypothèses multiples peut faire en sorte que le niveau de la procédure globale dépasse considérablement le niveau désiré. Alors que la plupart des méthodes d'inférence multiple sont conservatrices en présence de statistiques non indépendantes, les tests que nous proposons visent à contrôler exactement le niveau de signification. Pour ce faire, nous considérons des critères de test combinés proposés initialement pour des statistiques indépendantes. En appliquant la méthode des tests de Monte-Carlo, nous montrons comment ces méthodes de combinaison de tests peuvent s'appliquer à de tels cas, sans recours à des approximations asymptotiques. Après avoir passé en revue les résultats antérieurs sur ce sujet, nous montrons comment une telle méthodologie peut être utilisée pour construire des tests de normalité basés sur plusieurs moments pour les erreurs de modèles de régression linéaires. Pour ce problème, nous proposons une généralisation valide à distance finie du test asymptotique proposé par Kiefer et Salmon (1983) ainsi que des tests combinés suivant les méthodes de Tippett et de Pearson-Fisher. Nous observons empiriquement que les procédures de test corrigées par la méthode des tests de Monte-Carlo ne souffrent pas du problème de biais (ou sous-rejet) souvent rapporté dans cette littérature – notamment contre les lois *platikurtiques* – et permettent des gains sensibles de puissance par rapport aux méthodes combinées usuelles.

TESTS MULTIPLES SIMULÉS ET TESTS DE NORMALITÉ BASÉS SUR PLUSIEURS MOMENTS DANS LES MODÈLES DE RÉGRESSION*

Jean-Marie DUFOUR

Chaire de recherche du Canada en économétrie

Centre interuniversitaire de recherche en analyse des organisations (CIRANO)

Centre interuniversitaire de recherche en économie quantitative (CIREQ)

Département de sciences économiques

Université de Montréal

Abdeljelil FARHAT

CIRANO

CIREQ

Université de Montréal

Lynda KHALAF

CIREQ

Université de Montréal,

GREEN

Université Laval

RÉSUMÉ – Cet article illustre l'applicabilité des méthodes de rééchantillonnage dans le cadre des tests multiples (simultanés), pour divers problèmes économétriques. Les hypothèses simultanées sont une conséquence habituelle de la théorie économique, de sorte que le contrôle de la probabilité de rejet de combinaisons de tests est un problème que l'on rencontre fréquemment dans divers contextes économétriques et statistiques. À ce sujet, on sait que le fait d'ignorer le caractère conjoint des hypothèses multiples peut faire en sorte que le niveau de la procédure globale dépasse considérablement le niveau désiré. Alors que la plupart des méthodes d'inférence multiple sont conservatrices en présence de statistiques non indépendantes, les tests que nous proposons visent à contrôler exactement le niveau de signification. Pour ce faire, nous considérons des critères de test combinés proposés initialement pour des statistiques indépendantes. En appliquant la méthode des tests

* Cette recherche a bénéficié du support financier de la Chaire de recherche du Canada en économétrie, du Conseil des Arts du Canada (Bourse Killam), de l'Institut de Finance mathématique de Montréal (IFM2), du Conseil de recherche en sciences humaines du Canada, du Conseil de recherche en sciences naturelles et en génie du Canada, de la Fondation Alexander von Humboldt (Allemagne), du Réseau canadien de centres d'excellence (projet MITACS), du Fonds de recherche sur la société et la culture (Québec), et du Fonds de recherche sur la nature et les technologies (Québec).

de Monte-Carlo, nous montrons comment ces méthodes de combinaison de tests peuvent s'appliquer à de tels cas, sans recours à des approximations asymptotiques. Après avoir passé en revue les résultats antérieurs sur ce sujet, nous montrons comment une telle méthodologie peut être utilisée pour construire des tests de normalité basés sur plusieurs moments pour les erreurs de modèles de régression linéaires. Pour ce problème, nous proposons une généralisation valide à distance finie du test asymptotique proposé par Kiefer et Salmon (1983) ainsi que des tests combinés suivant les méthodes de Tippett et de Pearson-Fisher. Nous observons empiriquement que les procédures de test corrigées par la méthode des tests de Monte-Carlo ne souffrent pas du problème de biais (ou sous-rejet) souvent rapporté dans cette littérature – notamment contre les lois *platikurtiques* – et permettent des gains sensibles de puissance par rapport aux méthodes combinées usuelles.

ABSTRACT – This paper illustrates the usefulness of resampling-based methods in the context of multiple (simultaneous) tests, with emphasis on econometric applications. Economic theory often suggests joint (or simultaneous) hypotheses on econometric models; consequently, the problem of evaluating joint rejection probabilities arises frequently in econometrics and statistics. In this regard, it is well known that ignoring the joint nature of multiple hypotheses may lead to serious test size distortions. Whereas most available multiple test techniques are conservative in the presence of non-independent statistics, our proposed tests provably achieve size control. Specifically, we use the Monte-Carlo (MC) test technique to extend several well known combination methods to the non-independent statistics contexts. We first cast the multiple test problem into a unified statistical framework which: (i) serves to show how exact global size control is achieved through the MC test method, and (ii) yields a number of superior tests previously not considered. Secondly, we provide a review of relevant available results. Finally, we illustrate the applicability of our proposed procedure to the problem of moments-based normality tests. For this problem, we propose an exact variant of Kiefer and Salmon's (1983) test, and an alternative combination method which exploits the well known Fisher-Pearson procedure. Our simulation study reveals that the latter method seems to correct for the problem of test biases against platikurtic alternatives. In general, our results show that concrete and non-spurious power gains (over standard combination methods) can be achieved through our multiple Monte-Carlo test approach.

Marcel Dagenais a été l'un des pionniers de l'économétrie au Canada. Au Québec, il fut incontestablement le père de cette discipline et l'un des premiers économistes à publier des recherches de calibre international. Tout au long de sa carrière, il a combiné de manière judicieuse les travaux théoriques avec des recherches appliquées qui ont marqué l'introduction de méthodes statistiques rigoureuses dans l'étude de problèmes de politique économique. Sa contribution exceptionnelle à la science et à notre communauté a déjà été marquée de nombreuses manières, notamment par la première édition du Prix de la Société canadienne de science économique (lequel porte maintenant son nom), par le Prix Marcel-Vincent de l'Association canadienne française pour l'avancement des sciences, et par le Prix Léon-Gérin du Gouvernement du Québec. Au niveau personnel, ma dette envers lui est immense. C'est de lui que j'ai reçu mes premiers cours d'économétrie et qui m'a incité à poursuivre mes études dans ce domaine. C'est aussi lui qui m'a convaincu d'accepter un poste à l'Université de Montréal, après quoi j'ai eu le privilège de travailler avec lui comme coauteur sur des projets qui ont conduit à des publications dans des périodiques internationaux. Enfin, j'ai eu aussi eu l'honneur de recevoir à mon tour le Prix Marcel-Dagenais de

la Société canadienne de science économique. À tous les stades de ma carrière, il m'a appris mon métier. La liste de ceux qui ont bénéficié de son exemple, de ses conseils et de son leadership est très longue. Sa disparition fut pour moi une perte particulièrement grande. Cet article écrit avec deux autres chercheurs qui l'ont connu durant leurs études, Lynda Khalaf et Abdeljelil Farhat, est un modeste hommage à Marcel Dagenais.

Jean-Marie DUFOUR
Université de Montréal

INTRODUCTION

Combiner plusieurs tests non indépendants constitue un problème fréquent en économétrie et en statistique. En effet, les modèles économétriques suggèrent souvent de tester *plusieurs hypothèses* différentes ou encore de tester la *même hypothèse* au moyen de *plusieurs tests* différents, tout cela à partir des mêmes données. Le premier problème est typiquement associé à la construction de régions de confiance simultanées (*inférence simultanée*), tandis que le second conduit à combiner des tests différents qui sont valides sous la même hypothèse mais dont les caractéristiques de puissance varient suivant la contre-hypothèse considérée. Les tests obtenus en combinant plusieurs tests séparés constituent des *tests induits* (ou *tests combinés*). Bien que distincts, ces deux problèmes soulèvent des difficultés statistiques assez semblables et font intervenir les mêmes techniques. Le lecteur trouvera des discussions générales de ces problèmes dans Miller (1981), Folks (1984), Savin (1984), Dufour (1989) et Dufour et Torrès (1998).

Dans le cas de tests induits, le problème consiste à contrôler le niveau global de la procédure (c.-à-d., la probabilité de rejeter l'hypothèse nulle lorsque celle-ci est vraie) dans une situation où la distribution de chaque statistique de test est habituellement connue (ou relativement facile à obtenir) mais la loi conjointe des statistiques est inconnue (ou difficile à établir). En outre, il est facile de voir que l'utilisation de points critiques individuels, fournissant des tests au niveau désiré pour chaque statistique, peut conduire à un niveau global beaucoup plus élevé que celui des tests individuels. Par exemple, si le niveau de chaque test parmi 10 tests est égal à 0,05, la probabilité qu'*au moins un* de ces tests soit significatif est typiquement beaucoup plus élevée (pouvant aller jusqu'à 0,50). Il est donc impératif de tenir compte de la relation entre les différentes statistiques. Dans le cas de tests simultanés pour différentes hypothèses, le problème consiste à contrôler la probabilité de rejeter *au moins une hypothèse vraie* dans un ensemble d'hypothèses qui peut être considérable (dans certains cas, infini). Il est facile de voir que ce problème est une généralisation du précédent (où plusieurs hypothèses distinctes sont examinées plutôt qu'une seule) et, à nouveau, on doit tenir compte de la loi jointe des différentes statistiques. La possibilité de considérer plusieurs hypothèses devient importante lorsque l'on veut déterminer quel aspect d'une hypothèse jointe détermine le rejet de celle-ci, en vue par exemple d'améliorer la spécification d'un modèle. Les problèmes distributionnels soulevés par l'utilisation de plusieurs statistiques sont exceptionnellement complexes. Une méthode habituelle

dans ces contextes consiste à avoir recours à des *bornes* (par exemple, des bornes fondées sur des inégalités de type Boole-Bonferroni), mais celles-ci sont typiquement *conservatrices* et peuvent avoir un effet très négatif sur la puissance des procédures.

Récemment, les progrès de l'informatique ont rendu possible l'emploi de techniques de *rééchantillonnage* basées sur la simulation afin d'aborder des problèmes distributionnels complexes, qui ne se prêtent pas bien à la dérivation de la distribution exacte requise (pour une taille donnée d'échantillon) ni même parfois d'une approximation de grand échantillon. On distinguera notamment ici le *bootstrap* (voir, par exemple, Hall, 1992; Efron et Tibshirani, 1993; Davison et Hinkley, 1997; Horowitz, 1997; Chernick, 1999) et la méthode des tests de Monte-Carlo (Dwass, 1957; Barnard, 1963; Dufour et Khalaf, 2001; Dufour, 2002). Dans le cadre des tests d'hypothèses, ces méthodes sont, pour la plupart, proposées pour améliorer le contrôle de niveau dans les échantillons finis. De plus, pour certains problèmes non standards (par exemple, lorsque certains paramètres ne sont pas identifiés sous l'hypothèse nulle), les approches simulées sont non seulement plus fiables mais peuvent être plus faciles à mettre en oeuvre que les solutions asymptotiques disponibles; voir Dufour et Khalaf (2001), Dufour, Khalaf, Bernard et Genest (2003).

Les techniques de rééchantillonnage peuvent notamment être exploitées afin de résoudre divers problèmes de tests multiples. C'est le cas du *bootstrap*, qui permet d'améliorer de manière sensible la qualité des approximations asymptotiques; voir par exemple Westfall et Young (1993) et White (2000). L'emploi de résultats de convergence dans les échantillons finis reste toutefois sujet à caution¹. Afin d'obtenir des procédures dont la validité est démontrable à distance finie, nous avons récemment proposé d'appliquer la technique des tests de Monte-Carlo à différents problèmes économétriques; voir Dufour et Farhat (2001), Dufour et Khalaf (2002b, 2002c), Dufour, Khalaf, Bernard et Genest (2003) et Dufour, Khalaf et Beaulieu (2002, 2003). Spécifiquement, nous montrons comment, sans recourir à la méthode de Bonferroni, on peut obtenir des tests induits exacts sur divers modèles paramétriques, possiblement non gaussiens, et même non paramétriques. Pour ce faire, on calcule des seuils de signification marginaux (valeurs- p) combinés par l'application des tests de Monte-Carlo (Dufour, 2002). Les problèmes considérés incluent en particulier divers tests de spécification sur les résidus de régressions univariées et multivariées : corrélation contemporaine entre les résidus de plusieurs régressions, changement structurel, autocorrélation des erreurs, hétéroscédasticité, ainsi que des tests d'ajustement.

Dans le présent article, nous allons nous intéresser en priorité à la construction de tests induits. Plus précisément, nous poursuivons un double objectif. Premièrement, nous décrivons un cadre unifié qui permet de voir comment de nombreux

1. Pour des discussions générales du rôle et de la valeur des approximations asymptotiques en économétrie, voir Dufour (2001, 2003).

problèmes de *tests induits* peuvent être résolus en appliquant des techniques de tests de Monte-Carlo. Ceci permettra à la fois de réinterpréter des procédures déjà proposées et de considérer de nouvelles applications. Afin de mieux cerner le problème, nous réexaminons – à la lumière de ce cadre – deux exemples de tests induits dont le niveau est contrôlé par cette technique : (1) tests pour détecter une rupture de la variance des erreurs dans un modèle de régression; (2) tests d'indépendance entre les erreurs de plusieurs régressions linéaires (diagonalité de la matrice de covariance). Deuxièmement, nous proposons une application nouvelle à un problème économétrique important : tester la normalité des erreurs dans une régression en combinant plusieurs tests basés sur des moments différents, notamment des moments d'ordre supérieur à quatre. Alors que les tests habituels de ce type considèrent surtout les moments d'ordre trois et quatre – c.-à-d., les coefficients d'asymétrie et d'aplatissement – et sont fondés sur des approximations de grands échantillons (par exemple, Bowman et Shenton, 1975; Jarque et Bera, 1987), l'approche que nous décrivons ici fournit une méthode simple qui permet de combiner un nombre arbitraire de moments dans le cadre d'une procédure valide dans les échantillons finis. On obtient ainsi une version exacte du test de normalité proposé par Kiefer et Salmon (1983) ainsi que plusieurs modifications de celui-ci, qui peuvent fournir des gains de puissance.

Nous mettons l'emphasis sur le caractère général des procédures que nous proposons et qui s'appliquent à tout test multiple *simulable* sous l'hypothèse nulle. Pour ces problèmes, nous formulons un cadre général qui permet d'étendre l'application de tests combinés originalement proposés pour des statistiques indépendantes, spécifiquement les procédures suggérées par Tippett (1931) et Fisher-Pearson (Fisher, 1932; Pearson, 1933)². En appliquant la technique des tests de Monte-Carlo, nous généralisons ces procédures aux cas qui nous concernent.

Le plan de l'article est le suivant. Dans la première section, afin de motiver notre exposé, nous passons d'abord en revue divers problèmes pour lesquels notre méthodologie est pertinente. La deuxième section présente la méthodologie suggérée, tandis qu'à la troisième section nous décrivons deux applications spécifiques basées sur Dufour et Khalaf (2002b) et Dufour, Khalaf, Bernard et Genest (2003). Dans la quatrième section, nous décrivons une application originale à des tests de normalité des erreurs basés sur plusieurs moments dans un modèle de régression linéaire. Nous concluons à la dernière section.

1. TESTS MULTIPLES

Les tests de spécification constituent un des problèmes de base qui conduit à effectuer des tests multiples. Par exemple, les tests d'autocorrélation portent habituellement sur plusieurs retards et les tests d'ajustement distributionnel peuvent suggérer l'examen de moments à divers ordres. Dans ces cas, on applique typi-

2. Pour des discussions générales de ces procédures, le lecteur pourra consulter Folks (1984) et Dufour et Torrès (1998).

quement des tests *portmanteau* ou *omnibus* qui combinent des transformations asymptotiquement non corrélées des critères individuels, par exemple : (i) le test de normalité de JB (Jarque et Bera, 1980) qui combine un coefficient d'asymétrie avec un coefficient d'aplatissement ou celui de Kiefer et Salmon (1983) qui permet de considérer des moments d'ordre supérieur; (ii) le test portmanteau de Box et Pierce (1970) ou le test de ratio-de-variance multiple de Fong, Koh et Ouliaris (1997) contre la dépendance à plusieurs délais; ou (iii) le test de spécification conjoint de Bera-Jarque (Bera et Jarque, 1982) conçu pour tester simultanément l'absence de corrélation, l'homoscédasticité et la normalité. Ces tests peuvent avoir le bon niveau asymptotiquement, mais il est bien connu que le contrôle du niveau dans les échantillons finis peut être très insatisfaisant. En outre, on sait que le *rejet* d'une hypothèse jointe par un test omnibus peut être difficile à interpréter sans procéder à l'examen de plusieurs tests qui réagissent différemment à diverses contre-hypothèses.

Dans le cadre des tests de spécification, les tests de stabilité structurelle méritent une attention particulière. Le problème se pose lorsque le point de rupture est supposé inconnu. Par exemple, des tests de type Chow sont appliqués pour tout point de rupture éventuel, ce qui pose naturellement un problème de combinaison; voir Andrews (1993) et Andrews et Ploberger (1994). Cependant, la performance dans les échantillons finis des procédures de combinaison usuelles est insatisfaisante, ce qui motive la recherche d'autres méthodes, notamment celles basées sur des simulations. Pour des exemples pertinents, voir entre autres Christiano (1992), Diebold et Chen (1996) et Dufour et Kiviet (1996).

Les tests sur des modèles multivariés constituent un autre exemple typique. Les difficultés dans ce cas proviennent particulièrement de la *dimensionnalité*. Par exemple, on peut transformer un test d'hétéroscédasticité ou de normalité m -dimensionnel en une suite de m test univariés. Dans de tels cas, les tests omnibus sont couramment utilisés. Toutefois, ces tests reposent sur des approximations asymptotiques et peuvent très mal se comporter dans les échantillons finis; voir Bewley et Theil (1987), Zhou (1993), Deschamps (1996), Kilian et Demiroglu (2000), et Dufour, Khalaf et Beaulieu (2002, 2003).

2. TESTS MULTIPLES SIMULÉS À DISTANCE FINIE

Les divers exemples discutés dans la section précédente peuvent sembler *peu reliés*. Toutefois, malgré la diversité des hypothèses en question, tous ces problèmes ont un point en commun : on considère m statistiques S_i non indépendantes, chacune conçue pour tester une hypothèse nulle H_{0i} , $i = 1, \dots, m$ (où les H_{0i} peuvent être identiques). Afin de simplifier l'exposé (et sans perte de généralité), nous allons supposer que l'hypothèse H_{0i} est rejetée au niveau α lorsque S_i est « grand », c.-à-d. $S_i \geq c_i$ où c_i est un point critique tel que $P[S_i \geq c_i] \leq \alpha$ lorsque H_{0i} est « vraie ». De manière équivalente, le test $S_i \geq c_i$ peut être considéré comme significatif au niveau α lorsque $p_i \leq \alpha$ où p_i est le seuil de signification marginal du test (valeur- p). Dans le cas où S_i ne peut suivre qu'une seule distribution sous H_{0i} , on

peut écrire $p_i = G(S_i)$ où $G(x) = P[S_i \geq x]$ est la fonction de survie de S_i sous H_{0i} . Afin de simplifier l'exposé, nous supposons ici que chaque statistique S_i suit une loi continue sous H_{0i} . Dans ce cas, on voit aisément que p_i a une distribution uniforme sur l'intervalle $(0, 1)$ sous l'hypothèse nulle :

$$p_i \sim U(0, 1) \text{ sous } H_{0i}. \quad (1)$$

De façon plus concrète, nous formulons le problème qui nous intéresse ici comme suit : comment peut-on combiner ces différents tests pour évaluer l'hypothèse conjointe

$$H_0 : \text{« toutes les hypothèses } H_{0i} \text{ sont vraies »} \quad (2)$$

de façon à contrôler la probabilité de rejeter l'hypothèse jointe H_0 ?

Pour ce faire, nous proposons d'appliquer la technique de tests de Monte-Carlo (MC) que l'on peut résumer comme suit. En premier lieu, on obtient une statistique combinée que nous dénotons S . Afin de simplifier la présentation et sans perte de généralité, nous allons supposer que le test basé sur la statistique S rejette lorsque la statistique S est grande. Plusieurs choix sont possibles, incluant :

1. les tests de type Tippett (1931) fondés sur le minimum des valeurs- p ,

$$p_{\min} = \min_{i=1, \dots, m} \{p_i\} \quad (3)$$

ou encore

$$S_{\min} = 1 - \min_{i=1, \dots, m} \{p_i\}, \quad (4)$$

l'hypothèse H_0 est rejetée lorsque p_{\min} est petit (ou, de manière équivalente, lorsque S_{\min} est grand);

2. la procédure de Fisher-Pearson (Fisher, 1932; Pearson, 1933) qui considère le produit des valeurs- p ,

$$p_{\times} = \prod_{i=1}^m p_i \quad (5)$$

ou encore une des transformations suivantes de ce produit :

$$S_{\times} = 1 - \prod_{i=1}^m p_i, \quad (6)$$

$$S_{\ln} = -2 \sum_{i=1}^m \ln(p_i); \quad (7)$$

dans ce cas, on rejette H_0 lorsque p_{\times} est petit (de manière équivalente, lorsque S_{\times} (ou S_{\ln}) est grand).

Il est possible de considérer aussi un produit (ou une somme des logarithmes) pondéré

$$S_x^* = 1 - \prod_{i=1}^m p_i^{w_i}, \quad (8)$$

$$S_{\ln}^* = -2 \sum_{i=1}^m w_i \ln(p_i) \quad (9)$$

où les pondérations peuvent refléter des croyances *a priori* (Good, 1955) ou encore dépendre du vecteur des valeurs- p (Wilkinson, 1951). Pour un survol des différentes méthodes pour combiner plusieurs tests, le lecteur pourra consulter Folks (1984).

Considérons par exemple le cas des tests de ruptures structurelles. Tel que popularisé par Christiano (1992), évaluer conjointement les m critères de type Chow les plus grands peut apporter plus d'information que juste considérer leur supremum. Dans ce cas, m peut refléter une information *a priori* sur le nombre de ruptures multiples possibles. Nous pouvons donc formuler la suggestion de Christiano (1992) dans le cadre des statistiques combinées de la forme (8) en prenant $w_i = 1$ pour les q valeurs- p les plus petites et $w_i = 0$ pour les autres ($q < m$). L'exemple que nous traitons dans la section 3.1 illustre cette suggestion.

Dans le cas où les statistiques combinées que nous venons de décrire sont indépendantes, il est relativement facile de calculer leurs distributions sous l'hypothèse nulle. En effet, si les statistiques S_i sont indépendantes, les valeurs- p individuelles le sont aussi et on voit facilement que

$$\begin{aligned} P[p_{\min} \leq \alpha_0] &= 1 - P[p_1 > \alpha_0, \dots, p_m > \alpha_0] \\ &= 1 - \prod_{i=1}^m P[p_i > \alpha_0] = 1 - (1 - \alpha_0)^m. \end{aligned} \quad (10)$$

Il suffit alors de choisir $\alpha_0 = 1 - (1 - \alpha)^{1/m}$ pour que la région critique $p_{\min} \leq \alpha_0$ ait pour niveau α . De même, dans ce cas, on a

$$p_x \sim \prod_{i=1}^m U_i \quad \text{où} \quad U_1, \dots, U_m \stackrel{i.i.d.}{\sim} U(0, 1), \quad (11)$$

une distribution qui est facile à calculer (et simuler). On notera en particulier que

$$S_{\ln} = -2 \ln(p_x) \sim \chi^2(2m) \quad \text{sous } H_0, \quad (12)$$

de sorte que des points critiques pour p_x , S_x ou S_{\ln} peuvent être obtenus à partir de loi $\chi^2(2m)$. Malheureusement, quand les statistiques S_i ne sont pas indépendantes, ces résultats ne sont plus valides et calculer la distribution d'une statistique comme

p_{\min} ou p_x peut devenir extrêmement difficile. Toutefois, dans beaucoup de situations, il est facile de simuler cette distribution. On peut alors effectuer des tirages aléatoires dans la loi du modèle sous l'hypothèse nulle et construire les statistiques associées aux échantillons ainsi simulés.

Afin d'avoir une notation bien claire, nous désignerons par S_0 la statistique calculée à partir de l'échantillon observé et par S_1, \dots, S_N ses contreparties simulées. Finalement, on calcule une valeur- p MC basée sur le rang $\hat{R}_N(S_0)$ de la statistique observée dans la série S_0, S_1, \dots, S_N . La région critique du test MC correspond alors à

$$\hat{p}_N(S_0) = \frac{N\hat{G}_N(S_0; S_1, \dots, S_N) + 1}{N + 1} \leq \alpha \quad (13)$$

où $N\hat{G}_N(S_0; S_1, \dots, S_N)$ est le nombre de statistiques simulées qui sont plus grandes ou égales à S_0 . Par exemple, on rejette l'hypothèse nulle au seuil 5 % si la statistique observée S_0 excède le 95^e *quantile* des valeurs simulées.

La question qui se pose alors est : **sous quelles conditions cette procédure simulée contrôle-t-elle le niveau exact du test combiné?** Si la distribution sous l'hypothèse nulle de la statistique considérée est simulable et ne dépend d'aucun paramètre inconnu, on peut démontrer qu'une région critique de la forme (13) contrôle le niveau dans le sens où

$$P_{H_0}[\hat{p}_N(S_0) \leq \alpha] = \alpha \quad (14)$$

pourvu que $\alpha(N + 1)$ soit un nombre entier. La preuve repose sur des résultats distributionnels fondamentaux concernant la distribution des rangs associés à un ensemble fini de statistiques échangeables; voir Dufour et Khalaf (2001) et Dufour (2002).

Dans le cas qui nous concerne, une condition (suffisante) qui garantit le niveau d'un test de MC conjoint est une *pivotalité conjointe* : en d'autres termes, un test simulé tel que décrit ci-dessus satisfait (14) lorsque la distribution conjointe, sous l'hypothèse nulle (2), ne dépend pas de paramètres de nuisance; voir Dufour et Khalaf (2002a).

3. DEUX EXEMPLES ÉCONOMÉTRIQUES

Afin de rendre plus concrète la procédure générale décrite plus haut, nous allons examiner dans cette section deux exemples économétriques, où celle-ci a déjà été appliquée. Ces exemples portent sur les tests d'indépendance entre équations dans les modèles à régressions empilés (SURE, Dufour et Khalaf, 2002b), et des tests d'hétéroscédasticité contre une alternative de rupture de la variance en un point inconnu (Dufour, Khalaf, Bernard et Genest, 2003).

3.1 Tests d'indépendance dans les modèles SURE

Considérons un modèle SURE à m équations de la forme :

$$y_j = X_j \beta_j + u_j, j = 1, \dots, m \quad (15)$$

où y_j est un vecteur de T observations, X_j est une matrice (fixe ou strictement exogène) de plein rang colonne et de dimension $T \times k_j$, β_j comprend k_j coefficients inconnus et $u_j = (u_{1j}, u_{2j}, \dots, u_{Tj})'$ est un vecteur $T \times 1$ d'erreurs aléatoires. Soit $U_t = (u_{t1}, u_{t2}, \dots, u_{tm})'$ le vecteur des erreurs associé à la $t^{\text{ième}}$ observation. Concernant la loi des erreurs, nous allons supposer que

$$U_t = JW_t, t = 1, \dots, T \quad (16)$$

où J est une matrice triangulaire inconnue telle que

$$\Sigma \equiv J J' \text{ est non singulière} \quad (17)$$

et W_1, \dots, W_T sont des vecteurs aléatoires dont la distribution conjointe est complètement spécifiée (possiblement mais non nécessairement gaussienne). Nous voulons tester l'hypothèse que la matrice de corrélation contemporaine Σ est diagonale.

Pour ce problème, nous avons déjà montré que les statistiques usuelles (entre autres les critères LM et LR) sont *pivotales* (Dufour et Khalaf, 2002b). On peut alors obtenir facilement des versions exactes de ces tests sur la base de la méthode des tests de MC. Par ailleurs, Harvey et Phillips (1980, HP) ont proposé le test exact suivant afin de vérifier si le terme d'erreur d'une équation (disons la première) est indépendant des autres : on ajoute à la première régression les résidus correspondant aux autres équations et on applique un *test F* pour l'exclusion de ces résidus. Cette procédure est généralisée dans Dufour et Khalaf (2002b) afin de tester la forme diagonale de la matrice de covariance des erreurs. Deux types de tests induits sont proposés. Le premier (dénnoté F) combine les m critères de type HP obtenus en testant l'indépendance de chaque équation par rapport au reste du système. Le second test (dénnoté FS) examine consécutivement, si la première équation est indépendante des équations 2, ..., m , puis si la seconde est indépendante des équations 3, ..., m , et ainsi de suite. Pour chacun de ces deux ensembles de statistiques, on peut calculer des critères agrégés correspondants de type Tippet (dénnotés F_{\min} et F_x) et Fisher (FS_{\min} et FS_x).

Sous l'hypothèse nulle, la distribution conjointe des statistiques combinées par les critères F_{\min} , F_x , FS_{\min} et FS_x ne dépend pas de paramètres inconnus. En effet, on peut obtenir des répliques indépendantes dans la loi conjointe de ces statistiques sous l'hypothèse nulle, en effectuant des tirages dans la loi des W_t (qui est complètement spécifiée). Les variances des erreurs dans les différentes équations n'interviennent pas dans la distribution en question (sous l'hypothèse de diagonalité, les covariances sont nulles). La procédure de MC explicitée ci-dessus peut donc être appliquée facilement.

Le tableau 1 résume les résultats d'une étude de MC basée sur ce modèle avec $m = 5$ équations et 25 observations par équation. Ici LR réfère à la version MC du

test du quotient de vraisemblance. Dans le cas de tests à borne de type Bonferroni, les tests de type F sont obtenus en prenant $\alpha = \alpha / m$, tandis que, pour le test FS , nous avons utilisé une suite de niveaux qui décline suivant un schéma exponentiel (dont la somme est égale à α) : $\alpha_1 = \alpha / 2$, $\alpha_2 = \alpha / 2^2$, ..., $\alpha_{m-1} = \alpha / 2^{m-1}$, $\alpha_m = \alpha / 2^{m-1}$. La matrice Σ_1 a été obtenue à partir d'un exemple empirique, tandis que les autres matrices de covariance ont été construites en modifiant cette dernière de manière à réduire les covariances (en divisant les éléments de la décomposition Cholesky de Σ_1 par des constantes appropriées). Les éléments non nuls de Σ_0 correspondent à la diagonale de Σ_1 . Dans tous les cas, le niveau nominal des tests est 5 %, le nombre de réplifications est égal à 1 000, et les tests de MC utilisent 99 échantillons simulés³.

TABLEAU 1

PUISSANCE DES TESTS DE DIAGONALITÉ

| $T = 25$ | $\Sigma_0(H_0)$ | Σ_1 | Σ_2 | Σ_3 | Σ_4 |
|---|-----------------|------------|------------|------------|------------|
| Tests de MC (99 réplifications) | | | | | |
| LR | 0,047 | 1,0 | 0,980 | 0,824 | 0,494 |
| F_{\min} | 0,043 | 1,0 | 0,965 | 0,693 | 0,409 |
| F_x | 0,052 | 1,0 | 0,980 | 0,784 | 0,438 |
| FS_{\min} | 0,049 | 1,0 | 0,912 | 0,653 | 0,399 |
| FS_x | 0,052 | 1,0 | 0,984 | 0,799 | 0,562 |
| Tests de Harvey-Phillips généralisés (borne de Bonferroni) | | | | | |
| F_{\min} | 0,034 | 1,0 | 0,963 | 0,665 | 0,356 |
| FS_{\min} | 0,049 | 1,0 | 0,896 | 0,687 | 0,316 |

SOURCE : Dufour et Khalaf (2002b)

À partir de ces résultats, on voit que tous les tests MC contrôlent le niveau⁴. Les tests de MC induits ont une puissance nettement supérieure à celle des tests correspondants basés sur la borne de *Bonferroni*. La variante de Fisher du test de MC induit émerge comme étant le meilleur choix dans ce cas : sa puissance est excellente et surpasse même dans certains cas celle du test LR.

3. Des résultats plus détaillés sont disponibles dans Dufour et Khalaf (2002b).

4. Les niveaux empiriques des tests LR et LM standards (pour un seuil nominal de 5 %) sont 10,5 % et 26,7 % respectivement.

3.2 Tests de rupture d'une variance

Considérons maintenant une régression linéaire univariée :

$$y = X\beta + u \quad (18)$$

où $y = (y_1, \dots, y_T)'$, X est une matrice d'observations sur k régresseurs (fixes ou strictement exogènes), β est un vecteur inconnu de coefficients de régression, $u = (u_1, \dots, u_T)'$ est un vecteur d'erreurs tel que

$$u_t = \sigma_t \varepsilon_t, t = 1, \dots, T, \quad (19)$$

$\sigma_1, \dots, \sigma_T$ sont des paramètres d'échelle possiblement aléatoires et $\varepsilon = (\varepsilon_1, \dots, \varepsilon_T)'$ est un vecteur dont la distribution conditionnelle à X est complètement spécifiée (possiblement normale), par exemple

$$\varepsilon_1, \dots, \varepsilon_T \text{ sont i.i.d. suivant } F_0 \quad (20)$$

où F_0 est une loi donnée. Par exemple, nous pourrions considérer une distribution gaussienne sur les erreurs, c.-à-d.

$$\varepsilon_1, \dots, \varepsilon_T \stackrel{i.i.d.}{\sim} N(0, 1), \quad (21)$$

mais d'autres distributions pourraient être envisagées. L'hypothèse nulle qui nous intéressera ici est celle d'homogénéité (ou homoscedasticité) des erreurs :

$$H_0 : \sigma_t^2 = \sigma^2, t = 1, \dots, T, \quad (22)$$

ce qui implique que u_1, \dots, u_T sont i.i.d. selon la loi $P(u_t \leq v) = F_0(v / \sigma)$ sous H_0 .

Dans ce contexte, considérons le problème qui consiste à détecter des bris structuraux dans le paramètre d'échelle σ_t . Pour ce problème, Dufour, Khalaf, Bernard et Genest (2003) proposent des extensions des tests usuels, en particulier le test de Goldfeld et Quandt (1965, GQ), pour tenir compte d'un point de rupture non spécifié. Nous discutons cet exemple afin d'explicitier les critères de type (6).

Soit $GQ(T_1, T_3, k)$ le test de Goldfeld et Quandt (1965, GQ) usuel où k est le nombre de régresseurs (incluant la constante), T_1 et T_3 dénotent respectivement les tailles du premier et dernier sous-échantillon sur lequel le test est basé ($k < T_1$ et $k < T_3$), ce qui signifie que les $T_2 \equiv T - (T_1 + T_3)$ observations centrales sont omises. Sur la base d'une loi $F(T_3 - k, T_1 - k)$, on peut associer une valeur- p à $GQ(T_1, T_3, k)$, soit $pv(GQ; T_1, T_3, k)$. Pour tenir compte du point de rupture inconnu, nous considérons :

$$F_{\min}(GQ; K) \equiv 1 - \min_{(T_1, T_3) \in K} \{pv(GQ; T_1, T_3, k)\} \quad (23)$$

avec

$$K \equiv K(T, T_2) = \{(T_1, T_3) : L_0 \leq T_1 \leq U_0, T_3 = T - T_1 - T_2 \geq 0\}$$

où L_0 et U_0 sont les tailles minimales et maximales des sous-échantillons ($0 \leq T_2 \leq T - 2k - 2$, $L_0 \geq k + 1$, $U_0 \leq T - T_2 - k - 1$). En d'autres termes, cette définition

produit un ensemble de statistiques $\{GQ(T_1, T_3, k) : (T_1, T_3) \in K\}$ pour lesquelles le nombre d'observations centrales omis est maintenu constant. Nous considérons aussi

$$F_x(GQ; \hat{K}_{(4)}) \equiv 1 - \prod_{(T_1, T_3) \in \hat{K}_{(4)}} pv(GQ; T_1, T_3, k) \quad (24)$$

où $\hat{K}_{(4)}$ contient les quatre plus petites valeurs- p de l'ensemble $\{pv(GQ; T_1, T_3, k) : (T_1, T_3) \in K\}$. Le choix du nombre maximal des valeurs- p retenues (m dans ce cas) pourrait refléter le nombre de ruptures attendues ou encore le nombre de minima locaux dans la série $pv(GQ; T_1, T_3, k)$.

Sous l'hypothèse nulle, la distribution conjointe des statistiques combinées par les critères $F_{\min}(GQ; K)$ et $F_x(GQ; \hat{K}_{(4)})$ est *pivotale*. En effet, on peut obtenir des répliquations indépendantes dans la loi conjointe de ces statistiques sous l'hypothèse nulle d'homoscédasticité, en effectuant des tirages dans la loi des ε_t (qui est complètement spécifiée); voir Dufour, Khalaf, Bernard et Genest (2003). En d'autres termes, sous l'hypothèse nulle, le coefficient de régression β et le paramètre d'échelle σ^2 n'interviennent pas dans la distribution en question. La condition de validité des tests de MC est donc vérifiée.

Le tableau 2 résume les résultats d'une étude de MC qui compare le test GQ standard et les tests combinés définis ci-dessus⁵. Le modèle considéré comprend une équation de régression linéaire avec une taille d'échantillon $T = 50$, $k = 6$ régresseurs incluant une constante. La variance des erreurs pour l'observation t , $\sigma_t^2 = \sigma_1$, si $t \leq \tau_0$, et $\sigma_t^2 = \sigma_1 + \delta$, si $t > \tau_0$, où $\delta \geq 0$ et τ_0 représente la date de rupture (considérée inconnue). Les régresseurs sont tirés suivant une loi uniforme $U(0, 10)$ et les coefficients de régression sont égaux à un. De plus, $\alpha_0 = 1$, et (δ, τ_0) sont tels que : $(\sigma_1 + \delta) / \sigma_1 = 1, 4, 16$, et $\tau_0 / T = 0,3, 0,5, 0,7$. Le test GQ considéré est basé sur $T_2 = [T/5]$, $T_1 = T_3 = (T - T_2) / 2$. Pour chacun des critères combinés $F_{\min}(GQ; K)$ et $F_x(GQ; \hat{K}_{(4)})$, deux choix pour K sont considérés. Le premier est le plus grand possible $K^A = S_1(T, T_2, k + 1, T - T_2 - k - 1)$, où $T_2 = (T / 5)$. Le second est plus informatif, centré sur un intervalle prédéterminé autour de la vraie date de rupture $K^S = S_1(T, T_2, \tau_0^L(k), \tau_0^U(k))$, où $T_2 = (T / 5)$, $\tau_0^L(k) = \max\{k + 1, \tau_0 - [T / 5]\}$, $\tau_0^U(k) = \min\{T - k - T_2, \tau_0 + [T / 5]\}$. Dans tous les cas, le niveau nominal des tests est de 5 %, 1 000 répliquations sont considérées et les tests de MC utilisent 99 échantillons simulés.

Ces résultats illustrent clairement les avantages de considérer les tests combinés et surtout les critères de type Fisher-Pearson. En effet, les méthodes simultanées permettent d'atteindre une meilleure puissance et des décisions plus informatives.

5. Une description plus détaillée de ces résultats est disponible dans Dufour, Khalaf, Bernard et Genest (2003).

TABLEAU 2

PUISSANCE DES TESTS DE RUPTURE SUR LA VARIANCE,
AVEC POINTS DE RUPTURES NON SPÉCIFIÉS

| $T = 50$ | H_0 | $\tau_0 / T = 0,3$ | | $\tau_0 / T = 0,7$ | |
|---|-------|--------------------|------|--------------------|------|
| σ_2 / σ_1 | 1 | 2 | 4 | 2 | 4 |
| $GQ(F)$ | 5,5 | 35,0 | 59,9 | 57,1 | 99,3 |
| Test de MC combiné maximisé sur l'échantillon complet | | | | | |
| $F_x(GQ; K^A)$ | 5,6 | 56,8 | 98,6 | 73,3 | 99,7 |
| $F_{\min}(GQ; K^A)$ | 6,0 | 50,0 | 98,2 | 67,2 | 99,4 |
| Test de MC combiné maximisé sur un sous-échantillon | | | | | |
| $F_x(GQ; K^S)$ | 5,4 | 60,7 | 98,0 | 74,9 | 99,7 |
| $F_{\min}(GQ; K^S)$ | 6,2 | 53,1 | 98,3 | 75,0 | 99,5 |

SOURCE : Dufour et Khalaf-Bernard-Genest (2003)

4. TESTS DE NORMALITÉ BASÉS SUR PLUSIEURS MOMENTS

Nous allons maintenant considérer un problème pour lequel aucune procédure valide à distance finie n'a encore été proposée. Il s'agit de tester la normalité des erreurs dans un modèle de régression à partir d'un nombre arbitraire de moments empiriques des résidus. En particulier, on voudrait pouvoir considérer des moments d'ordre supérieur au quatrième moment. Le modèle considéré est la régression linéaire (18) avec $\sigma_t^2 = \sigma^2$, $t = 1, \dots, T$. Dans ce modèle, les tests de normalité sont typiquement basés sur le vecteur des résidus des moindres carrés ordinaires :

$$\hat{u} = (\hat{u}_1, \dots, \hat{u}_T)' = y - X \hat{\beta} = M_X u \tag{25}$$

où $\hat{\beta} = (X' X)^{-1} X' y$ et $M_X = I_T - X(X' X)^{-1} X'$. Soit

$$s^2 = (T - k)^{-1} \sum_{i=1}^T \hat{u}_i^2, \quad \hat{\sigma}^2 = T^{-1} \sum_{i=1}^T \hat{u}_i^2. \tag{26}$$

Nous supposons en outre ici que la matrice X comprend le régresseur constant $\iota_T = (1, \dots, 1)'$. Lorsque $X = \iota_T$, nous obtenons le modèle moyenne-variance usuel, dans le cadre duquel la majorité des tests de normalité est étudiée en statistique.

Le problème consiste à tester

$$H_0 : f(u) = \varphi(0, \sigma^2) \quad (27)$$

où $f(u)$ est la fonction de densité inconnue et $\varphi(\mu, \sigma^2)$ représente la densité normale centrée de variance σ^2 . Cette question constitue un problème fondamental en statistique et en économétrie. En effet, dans le cadre des tests d'ajustement, l'hypothèse de normalité joue un rôle particulièrement important, étant donné la popularité de cette distribution tant sur le plan théorique que sur le plan pratique; pour des survols de la littérature pertinente, voir Mardia (1980), D'Agostino et Stephens (1986), Dufour, Farhat, Gardiol et Khalaf (1998) et Thode (2002). Évidemment, un très grand éventail de tests de normalité est disponible; par exemple, Thode (2002) recense au moins 40 tests différents sans compter les extensions et variantes de ceux-ci.

Dans ce contexte, nous avons déjà montré que la méthode des tests de MC permet de résoudre le problème du contrôle de niveau, pour la classe des tests de normalité invariants à la localisation et à l'échelle (qui contient presque sans exception toutes les procédures couramment utilisées), voir Dufour *et al.* (1998). Alors que les distributions de ces statistiques (à distance finie et même asymptotiquement) sont complexes et habituellement spéciales, les valeurs- p exactes obtenues par MC demeurent très faciles à calculer. Cette flexibilité permet d'envisager de nouveaux tests dont les caractéristiques de puissance seraient plus prometteuses. Il est important de rappeler que le problème du contrôle de niveau (ou concrètement la disponibilité d'une théorie distributionnelle exacte ou asymptotique) pose des limites réelles sur le choix des statistiques. Dans le cas qui nous concerne, à la lumière des résultats dans Dufour *et al.* (1998), la seule contrainte dont nous devons tenir compte consiste à construire les statistiques de test en fonction du vecteur des résidus standardisés, ce qui n'est guère restrictif en pratique.

Ce résultat nous incite à étudier d'autres procédures de test, notamment des procédures qui considèrent plusieurs moments. En effet, une famille couramment appliquée de tests de normalité s'appuie sur les déviations entre les moments empiriques des résidus et ceux (connus) de la loi normale. Même si les tests usuels se limitent aux troisième et quatrième moment (voir Jarque et Bera, 1987; Dufour *et al.*, 1998), Kiefer et Salmon (1983) ont proposé d'exploiter les moments d'ordre supérieur à quatre. Dans notre cas, ceci conduit à considérer les moments empiriques

$$\hat{\mu}_j = \frac{1}{T} \sum_{i=1}^T (\hat{u}_i / s)^j, \quad j = 3, \dots, 7,$$

ainsi que les transformations suivantes de ces moments :

$$M_3 = T \hat{\mu}_3 / 6,$$

$$M_4 = T(\hat{\mu}_4 - 3)^2 / 24,$$

$$M_5 = T(\hat{\mu}_5 - 10\hat{\mu}_3)^2 / 120,$$

$$M_6 = T(\hat{\mu}_6 - 15\hat{\mu}_4 + 30)^2 / 720,$$

$$M_7 = T(\hat{\mu}_7 - 21\hat{\mu}_5 + 105\hat{\mu}_3)^2 / 5\,040.$$

Sous l'hypothèse de normalité des erreurs, et dans le cadre du modèle moyenne-variance, Kiefer et Salmon (1983) démontrent que chacun des ces critères suit asymptotiquement une distribution $\chi^2(1)$; de plus, ces critères sont asymptotiquement orthogonaux, ce qui permet de définir des tests portmanteau basés sur leurs sommes⁶. Par exemple, le test le plus couramment utilisé en économétrie, qui est dû à Jarque et Bera (1987), combine les coefficients d'asymétrie et d'aplatissement (basés sur les troisième et quatrième moments) :

$$JB = \sum_{j=3}^4 M_j. \quad (28)$$

Suite aux arguments de Kiefer et Salmon (1983), on peut aussi considérer une statistique du type

$$S(3, \bar{m}) = \sum_{j=3}^{\bar{m}} M_j. \quad (29)$$

où \bar{m} est l'ordre maximal considéré. Par exemple, Kiefer et Salmon (1983) définissent la statistique $S(3, 5)$. Dans ce qui suit, nous allons prendre $\bar{m} = 7$, c.-à-d.

$$S(3, 7) = \sum_{j=3}^7 M_j. \quad (30)$$

Il est important de rappeler que ces moments empiriques, particulièrement les coefficients d'asymétrie et d'aplatissement, sont largement étudiés dans la littérature. Pour une revue historique concernant les coefficients d'asymétrie et d'aplatissement, qui remontent aux travaux de Fisher (1930), et un survol des résultats concernant leurs distributions, le lecteur pourra consulter D'Agostino et Stephens (1986) et Thode (2002). Campbell, Lo et MacKinlay (1997 : 18-20) proposent une approximation asymptotique ($N(0, 6/T)$ et $N(0, 24/T)$ respectivement) pour leurs lois sous une hypothèse nulle compatible avec des données financières.

Les résultats de l'étude de puissance rapportée par Dufour *et al.* (1998) illustrent la supériorité des procédures basées sur les troisième et quatrième moments, une fois leur niveau sous contrôle. Les études résumées dans Thode (2002) semblent corroborer cette évidence. Nous proposons ici des versions exactes des tests de Kiefer et Salmon (1983), ainsi que des statistiques modifiées en combinant les différents moments suivant les arguments de la section 2. Ainsi, notre procédure de combinaison permet de considérer :

6. Pour une généralisation de cette approche à un cadre GMM, le lecteur peut aussi consulter Bontemps et Meddahi (2005).

$$P(3, 7) = 1 - \prod_{i=3}^7 p_i, \quad (31)$$

$$\text{Min}(3, 7) = 1 - \min_{3 \leq i \leq 7} p_i \quad (32)$$

où p_j = la valeur- p associée à M_j (par exemple, avec des points critiques approximatifs découlant de la loi $\chi^2(1)$).

Tel que démontré dans Dufour *et al.* (1998), les moments standardisés des résidus sont invariants par rapport au coefficient de régression et au paramètre d'échelle σ^2 . En d'autres termes, on peut obtenir des valeurs simulées pour ces moments sous l'hypothèse de normalité des erreurs en effectuant des tirages dans la loi normale standard, conditionnelle à la matrice de projection M_x . Il s'ensuit que la distribution conjointe des moments M_i , $i = 1, \dots, 7$ est *pivotal* sous l'hypothèse nulle conjointe. Un test de MC basé sur (7)-(8) conduit à des procédures différentes de celles de Kiefer et Salmon (1983). Il est donc intéressant d'examiner leur performance vis-à-vis des tests asymptotiques et des versions exactes de ces derniers.

Comme cette question n'a pas été étudiée dans le cadre d'un modèle moyenne-variance, nous nous concentrerons sur ce modèle simple pour évaluer, par simulation, la performance relative des différentes procédures. Nous considérons une taille d'échantillon de 25 observations. À lumière des résultats dans Dufour *et al.* (1998), il est important de considérer une taille qui n'est pas trop grande afin de comparer la puissance des différents tests de normalité⁷. Les observations sont simulées sous la forme $M_x u$; il n'est donc pas nécessaire de spécifier une valeur pour β et, sans perte de généralité, on peut poser $\sigma^2 = 1$.

Pour chaque échantillon (de taille $T = 25$) nous avons calculé la moyenne et la variance empirique, les statistiques M_i , $i = 1, \dots, 7$ ainsi que les statistiques port-manteau JB , $S(3, 7)$, $P(3, 7)$ et $\text{Min}(3, 7)$. Nous rapportons, dans le tableau 3, les rejets empiriques (en pourcentage, sur 10 000 répliquions) sous l'hypothèse nulle et diverses lois alternatives. Dans ce tableau, \mathcal{N} , \mathcal{B} , \mathcal{C} , Γ , \mathcal{L}_T et $t(5)$ dénotent respectivement la loi normale, la loi Bêta(2, 3), la loi de Cauchy, la loi Gamma(2, 1), la loi log-normale centrée réduite et la loi de Student $t(5)$. La partie du tableau intitulée « tests standards » se base sur les développements asymptotiques de Kiefer et Salmon (1983). Nos résultats se résument comme suit.

1. Les tests standards sous-rejettent sérieusement. Les conséquences de ce sous-rejet sur la puissance des tests sont évidentes. La technique de MC améliore ainsi la puissance effective des tests.

7. La puissance des tests basés sur les moments atteint 1 très rapidement (dès $T = 50$). Bien entendu, la vitesse de convergence peut être influencée par la structure des régresseurs.

2. Dans le cas de la loi Bêta, les tests omnibus de type somme (à la Kiefer et Salmon, 1983) sont tous *biaisés*. La méthode de MC ne corrige pas ce problème, dans le cas du test de Jarque-Bera, même avec l'ajout de moments supérieurs. Par contre, malgré le fait que la puissance des tests induits $P(3, 7)$ et $\text{Min}(3, 7)$ reste faible, elle est supérieure au niveau. Notons que tous les tests étudiés dans Dufour *et al.* (1998), notamment le test de Jarque-Bera, sont biaisés dans ce cas.
3. La version de MC du test JB fournit la meilleure puissance dans le cas de la loi de Student $t(5)$. Par contre, pour la loi Gamma(2, 1), l'ajout des moments supérieurs semble pertinent, et une certaine amélioration de la puissance est observable pour les lois log-normale et de Cauchy, par rapport au test JB. Enfin, les tests $P(3, 7)$ et $\text{Min}(3, 7)$ sont préférables aux tests de type somme pour les alternatives Bêta et Cauchy.

TABLEAU 3

PUISSANCE DES TESTS DE NORMALITÉ BASÉS SUR LES MOMENTS

| | Tests standards | | | | | | Tests de MC | | | | | |
|--------------------|-----------------|---------------|------|----------|-----------------|--------|---------------|---------------|------|----------|-----------------|--------|
| | \mathcal{N} | \mathcal{B} | C | Γ | \mathcal{L}_T | $t(5)$ | \mathcal{N} | \mathcal{B} | C | Γ | \mathcal{L}_T | $t(5)$ |
| M_3 | 3,2 | 1,7 | 79,7 | 49,7 | 98,5 | 19,3 | 5,0 | 3,3 | 81,1 | 56,1 | 98,8 | 22,3 |
| M_4 | 1,7 | 0,1 | 88,0 | 20,9 | 82,9 | 17,8 | 5,0 | 5,5 | 91,8 | 29,4 | 87,9 | 25,4 |
| M_5 | 0,6 | 0,7 | 69,4 | 15,5 | 68,8 | 8,6 | 5,2 | 9,2 | 76,0 | 42,8 | 79,6 | 15,5 |
| M_6 | 0,2 | 0,1 | 68,4 | 5,9 | 56,5 | 5,8 | 5,2 | 9,4 | 80,5 | 13,6 | 69,9 | 16,0 |
| M_7 | 0,3 | 0,1 | 53,8 | 3,2 | 50,6 | 3,6 | 5,0 | 4,3 | 63,2 | 13,9 | 60,0 | 14,3 |
| JB | 2,7 | 0,7 | 89,7 | 37,6 | 95,8 | 21,3 | 4,8 | 2,1 | 91,4 | 48,0 | 97,7 | 26,0 |
| $S(3, 7)$ | 1,9 | 0,6 | 90,1 | 32,6 | 96,7 | 18,3 | 4,9 | 4,4 | 91,0 | 56,2 | 99,2 | 24,7 |
| $P(3, 7)$ | — | — | — | — | — | — | 4,9 | 5,8 | 93,1 | 43,3 | 97,6 | 25,4 |
| $\text{Min}(3, 7)$ | — | — | — | — | — | — | 5,0 | 6,0 | 93,3 | 48,3 | 98,8 | 24,4 |

On voit, par ces exemples, qu'il est possible de réaliser des gains de puissance par rapport aux méthodes de combinaison usuelles. De plus, il semble que les tests combinés selon Tippett ou Fisher-Pearson ne souffrent pas du problème de biais (contre les lois *platikurtiques*) souvent rapporté dans cette littérature. Nous estimons donc que cette expérience motive des travaux plus poussés sur ce sujet,

notamment dans un contexte multivarié. Dans ce cas, le problème (supplémentaire) consiste à combiner les critères de tests entre équations voir là-dessus Kilian et Demiroglu (2000) et Dufour, Khalaf et Beaulieu (2003).

CONCLUSION

Cet article décrit des stratégies de tests multiples fondées sur des simulations et dont la validité peut être démontrée pour une grande classe de modèles économétriques. En particulier, nous avons décrit un cadre unifié qui permet de montrer comment de nombreux problèmes de *tests induits* peuvent être résolus en appliquant des techniques de tests de Monte-Carlo. Ceci permet en particulier de voir comment les méthodes classiques proposées par Tippett (1931) et Fisher-Pearson (Fisher, 1932; Pearson, 1933) pour combiner des tests indépendants s'étendent à des statistiques non indépendantes. Ensuite, après avoir montré comment certaines procédures proposées dans Dufour et Khalaf (2002b) et Dufour, Khalaf, Bernard et Genest (2004) peuvent s'interpréter comme des cas spéciaux de l'approche considérée, nous avons proposé une application nouvelle qui permet d'obtenir des tests de normalité pour les résidus d'une régression basés sur un nombre quelconque de moments. On obtient ainsi une version exacte de la procédure de Kiefer et Salmon (1983) ainsi que des tests combinés de types Tippett et Fisher-Pearson. Les résultats que nous rapportons pour les exemples étudiés suggèrent que les tests de MC induits présentent des avantages par rapport aux méthodes courantes.

Nous soulignons le caractère général des procédures que nous proposons et que ces dernières peuvent s'étendre à toute espèce de tests multiples lorsque la *loi conjointe* des statistiques de test est *simulable* sous l'hypothèse nulle. Par conséquent, des raffinements et des extensions selon les orientations évoquées plus haut, permettent d'anticiper des résultats fort utiles en économétrie du point de vue théorique et empirique.

BIBLIOGRAPHIE

- ANDREWS, D. W. K. (1993), « Tests for Parameter Instability and Structural Change with Unknown Change Point », *Econometrica*, 61 : 821-856.
- ANDREWS, D. W. K. et W. PLOBERGER (1994), « Optimal Tests when a Nuisance Parameter is Present only under the Alternative », *Econometrica*, 62 : 1 383-1 414.
- BARNARD, G. A. (1963), « Comment on 'The Spectral Analysis of Point Processes' by M. S. Bartlett », *Journal of the Royal Statistical Society, Series B*, 25 : 294.
- BERA, A. K. et C. M. JARQUE (1982), « Model Specification Tests: A Simultaneous Approach », *Journal of Econometrics*, 20 : 59-82.
- BEWLEY, R. et H. THEIL (1987), « Monte-Carlo Testing for Heteroscedasticity in Equation Systems », *Advances in Econometrics*, 6 : 1-15.

- BONTEMPS, C. et N. MEDDAHI (2005), « Testing Normality: A GMM Approach », *Journal of Econometrics*, 124 : 149-186.
- BOWMAN, K. O. et B. R. SHENTON (1975), « Omnibus Test Contours for Departures from Normality Based on $\sqrt{b_1}$ and b_2 », *Biometrika*, 52 : 591-611.
- BOX, G. E. P. et D. A. PIERCE (1970), « Distribution of Residual Autocorrelations in Autoregressive-integrated Moving Average Time Series Models », *Journal of the American Statistical Association*, 65 : 1 509-1 526.
- CAMPBELL, Y. Y., A. W. LO et A. C. MACKINLAY (1997), *The Econometrics of Financial Markets*, Princeton University Press, New Jersey.
- CHERNICK, M. R. (1999), *Bootstrap Methods: A Practitioner's Guide*, John Wiley & Sons, New York.
- CHRISTIANO, L. J. (1992), « Searching for a Break in GNP », *Journal of Business and Economic Statistics*, 10 : 237-249.
- D'AGOSTINO, R. B. et M. A. STEPHENS (éds) (1986), *Goodness-of-Fit Techniques*, Marcel Dekker, New York.
- DAVISON, A. et D. HINKLEY (1997), *Bootstrap Methods and Their Application*, Cambridge University Press, Cambridge (UK).
- DESCHAMPS, P. (1996), « Monte-Carlo Methodology for LM and LR Autocorrelation Tests in Multivariate Regressions », *Annales d'Économie et de Statistique*, 43 : 150-169.
- DIEBOLD, F. X. et C. CHEN (1996), « Testing Structural Stability with Endogenous Break Point: A Size Comparison of Analytic and Bootstrap Procedures », *Journal of Econometrics*, 70 : 221-241.
- DUFOUR, J.-M. (1989), « Nonlinear Hypotheses, Inequality Restrictions, and Non-nested Hypotheses: Exact Simultaneous Tests in Linear Regressions », *Econometrica*, 57 : 335-355.
- DUFOUR, J.-M. (2001), « Logique et tests d'hypothèses : réflexions sur les problèmes mal posés en économétrie », *L'Actualité économique*, 77(2) : 171-190.
- DUFOUR, J.-M. (2002), « Monte-Carlo Tests with Nuisance Parameters: A General Approach to Finite-sample Inference and Nonstandard Asymptotics in Econometrics », *Journal of Econometrics*, à paraître.
- DUFOUR, J.-M. (2003), « Identification, Weak Instruments and Statistical Inference in Econometrics », *Canadian Journal of Economics*, 36(4) : 767-808.
- DUFOUR, J.-M. et A. FARHAT (2001), « Exact Nonparametric Two-sample Homogeneity Tests for Possibly Discrete Distributions », Technical report, C.R.D.E., Université de Montréal, 26 pages.
- DUFOUR, J.-M., A. FARHAT, L. GARDIOL et L. KHALAF (1998), « Simulation-based Finite Sample Normality Tests in Linear Regressions », *The Econometrics Journal*, 1 : 154-173.
- DUFOUR, J.-M. et L. KHALAF (2001), « Monte-Carlo Test Methods in Econometrics », in B. BALTAGI (éd.), *Companion to Theoretical Econometrics*, Blackwell Companions to Contemporary Economics, Basil Blackwell, Oxford, U.K., chapter 23, p. 494-519.

- DUFOUR, J.-M. et L. KHALAF (2002a), « Exact Simulation Based Multiple Hypothesis Tests », Technical report, CIRANO et CIREQ, Université de Montréal et Département d'économique, Université Laval.
- DUFOUR, J.-M. et L. KHALAF (2002b), « Exact Tests for Contemporaneous Correlation of Disturbances in Seemingly Unrelated Regressions », *Journal of Econometrics*, 106(1) : 143-170.
- DUFOUR, J.-M. et L. KHALAF (2002c), « Simulation Based Finite and Large Sample Tests in Multivariate Regressions », *Journal of Econometrics*, 111(2) : 303-322.
- DUFOUR, J.-M., L. KHALAF et M.-C. BEAULIEU (2002), « Finite-sample Diagnostics for Multivariate Regressions with Applications to Linear Asset Pricing Models », Technical report, CIRANO et CIREQ, Université de Montréal et Département d'économique, Université Laval.
- DUFOUR, J.-M., L. KHALAF et M.-C. BEAULIEU (2003), « Exact Skewness-kurtosis Tests for Multivariate Normality and Goodness-of-fit in Multivariate Regressions with Application to Asset Pricing Models », *Oxford Bulletin of Economics and Statistics*, 65 : 891-906.
- DUFOUR, J.-M., L. KHALAF, J.-T. BERNARD et I. GENEST (2004), « Simulation-based Finite-sample Tests for Heteroskedasticity and ARCH Effects », *Journal of Econometrics*, 122 : 319-347.
- DUFOUR, J.-M. et J. F. KIVIET (1996), « Exact Tests for Structural Change in First-order Dynamic Models », *Journal of Econometrics*, 70 : 39-68.
- DUFOUR, J.-M. et O. TORRÈS (1998), « Union-intersection and Sample-split Methods in Econometrics with Applications to SURE and MA Models », in D. E. A. GILES et A. ULLAH (éds), *Handbook of Applied Economic Statistics*, Marcel Dekker, New York, p. 465-505.
- DWASS, M. (1957), « Modified Randomization Tests for Nonparametric Hypotheses », *Annals of Mathematical Statistics*, 28 : 181-187.
- EFRON, B. et R. J. TIBSHIRANI (1993), *An Introduction to the Bootstrap*, vol. 57, *Monographs on Statistics and Applied Probability*, Chapman & Hall, New York.
- FISHER, R. A. (1930), « The Moments of the Distribution for Normal Samples of Measures for Departures from Normality », *Proceedings of the Royal Society of London A*, 130 : 16-28.
- FISHER, R. A. (1932), *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh.
- FOLKS, J. L. (1984), « Combination of Independent Tests », in P. R. KRISHNAIAH et P. K. SEN (éds), *Handbook of Statistics 4: Nonparametric Methods*, North-Holland, Amsterdam, p. 113-121.
- FONG, M. F., S. K. KOH et S. OULIARIS (1997), « Joint Variance-ratio Tests of the Martingale Hypothesis for Exchange Rates », *Journal of Business and Economic Statistics*, 15 : 51-59.
- GOLDFELD, S. M. et R. QUANDT (1965), « Some Tests for Heteroscedasticity », *Journal of the American Statistical Association*, 60 : 539-547.

- GOOD, I. J. (1955), « On the Weighted Combination of Significance Tests », *Journal of the Royal Statistical Society, Series B*, 17 : 264-265.
- HALL, P. (1992), *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, New York.
- HARVEY, A. C. et G. D. A. PHILLIPS (1980), « Testing for Serial Correlation in Simultaneous Equation Models », *Econometrica*, 48 : 747-759.
- HOROWITZ, J. L. (1997), « Bootstrap Methods in Econometrics: Theory and Numerical Performance », in D. KREPS et K.W. WALLIS (éds), *Advances in Economics and Econometrics*, vol. 3, Cambridge University Press, Cambridge, U.K., p. 188-222.
- JARQUE, C. M. et A. K. BERA (1980), « Efficient Tests for Normality, Heteroscedasticity and Serial Independence of Regression Residuals », *Economics Letters*, 6 : 255-259.
- JARQUE, C. M. et A. K. BERA (1987), « A Test for Normality of Observations and Regression Residuals », *International Statistical Review*, 55 : 163-172.
- KIEFER, N. M. et M. SALMON (1983), « Testing Normality in Econometric Models », *Economic Letters*, 11 : 123-127.
- KILIAN, L. et U. DEMIROGLU (2000), « Residual-based Tests for Normality in Autoregressions: Asymptotic Theory and Simulation Evidence », *Journal of Business and Economic Statistics*, 18 : 40-50.
- MARDIA, K. V. (1980), « Tests of Univariate and Multivariate Normality », in P. R. KRISHNAIAH (éd.), *Handbook of Statistics I: Analysis of Variance*, North-Holland, Amsterdam, p. 279-320.
- MILLER, Jr., R. G. (1981), *Simultaneous Statistical Inference*, 2^e éd., Springer-Verlag, New York.
- PEARSON, K. (1933), « On a Method of Determining Whether a Sample of Size n Supposed to Have Been Drawn from a Parent Population », *Biometrika*, 25 : 379-410.
- SAVIN, N. E. (1984), « Multiple Hypothesis Testing », in Z. GRILICHES et M. D. INTRILLIGATOR (éds), *Handbook of Econometrics, Volume 2*, North-Holland, Amsterdam, chap. 14, p. 827-879.
- THODE, H. C. Jr., (2002), *Testing for Normality*, number 164 in *Statistics: Textbooks and Monographs*, Marcel Dekker, New York.
- TIPPETT, L. H. (1931), *The Methods of Statistics*, Williams and Norgate, London.
- WESTFALL, P. H. et S. S. YOUNG (1993), *Resampling-Based Multiple Testing: Examples and Methods for p -Value Adjustment*, John Wiley & Sons, New York.
- WHITE, H. (2000), « A Reality Check for Data Snooping », *Econometrica*, 68 : 1 097-1 126.
- WILKINSON, B. (1951), « A Statistical Consideration in Psychological Research », *Psychology Bulletin*, 48 : 156-158.
- ZHOU, G. (1993), « Asset-pricing Tests Under Alternative Distributions », *The Journal of Finance*, 48 : 1 927-1 942.